

# **THE ROLE OF INNOVATION AND STATISTICAL METHODOLOGY IN SAFETY ASSESSMENT PROJECTS**

**Gerald R. Stewart**  
**National Highway Traffic Safety Administration**  
**United States**  
**Paper # 412**

## **ABSTRACT**

This paper discusses safety assessment projects which are characterized by availability of advanced vehicle information technologies. Present and future crash event recording situations are discussed to illustrate the use of innovation and statistical analysis of appropriate data for safety evaluations.

Present development, testing and implementation of in-vehicle event recording devices are discussed in terms of new and emerging technologies and the innovations needed to promote improved data quality and ease of use in safety assessments. Approaches to evaluating statistical properties of new data sources are also discussed.

Present and future computer technology and software, event recording devices and data for analysis are advancing rapidly but still there is a need for innovation and study of statistical properties of data sources.

## **SENSING TECHNOLOGIES**

An example application of high speed sensing is the case where occupants of motor vehicles are protected from fatal injury in a crash when air-bags are deployed by sensors which detect and/or predict an above threshold crash pulse which is known to be at a level that would most likely result in a fatal injury to the occupant. Some vehicle crashes are serious and the physical conditions which signal the seriousness are very complex (see pages 30-38 of [1]).

In addition to air-bags, other methods are being developed and tested to help save lives in crash situations. Crash event recorders have the primary function of triggering the vehicle's protective devices such as air-bags but other options are being evaluated also. One option is to provide for automatic notification to medical emergency providers in the case of a serious crash so that appropriate medical treatment can be provided at the right time to attend to the crash trauma in a life saving manner [2,3].

Another option for crash event data recording equipment is the preservation of important data that can be used to evaluate a variety of crash situations at some future date. This type of data provides better

opportunities to focus on aspects of crashes and/or vehicles that can be used to make decisions which will help both vehicles, their occupants and the roadway to operate at safer levels than before.

Analysis [2] has shown that for serious crashes, the use of appropriate emergency medical services at the right times can be a significant factor in the outcome of injuries to victims of crashes. Crash victims who are suffering life threatening trauma must be given treatment to keep their airway open and their critical body functions going long enough to recover from the impact of the crash. Event data recording technology now exists to provide an in-vehicle device that can detect a serious crash and send a cell phone call to appropriate authorities [3].

Other technologies [4,5] provide digital images from on-vehicle video cameras which can be sent from the crash scene electronically to information processing stations on the Internet. This form of data can also be useful in post-crash analysis of the circumstances of the crash.

## **THE USE OF VIN STRUCTURE FOR RECORDED DATA**

Innovative ideas can be used to promote data structures which include items from these new and emerging technologies for in-vehicle event recorders. For example, one important data item to consider is the VIN (Vehicle Identification Numbers) of the vehicle which experiences an event of interest.

Vehicle Identification Number Standardization has been established for all passenger cars, multi-purpose passenger vehicles, trucks, buses, trailers, incomplete vehicles and motorcycles for a model years beginning with model year 1981.

The first three characters of the VIN are designated the WMI (World Manufacturer Identification). The WMI uniquely identifies the Nation of Origin, Manufacturer, Make and Type of Vehicle. When a 9 appears in position 3 of the VIN, it signifies small manufacturers which produce less than 1000 vehicles per year. For these cases, positions 12-14 of the VIN are needed to identify the small manufacturer.

The second section has five characters and has been designated the VDS (Vehicle Description Section). The VDS uniquely identifies the attributes of the vehicle such as Model, Body Style, Engine etc. The third section of the VIN starts in position 10 and goes to position 17 with position 11 designating the plant where the vehicle was manufactured. The Vin characters in position 12-17 are of special interest because of their linkage to the registration information which can be used to identify the owner of the vehicle. The character in position 9 of the VIN is a check digit which is calculated from the other 16 characters of the vin. The check digit is used to verify the validity of a VIN. Vin position 10 is used to record the model year of the vehicle.

VIN interpretation software can be used to identify vehicles and to link to a variety of information on vehicle characteristics such as wheelbase, engine and transmission , restraint system, brake design, anti-theft devices , weight-to-horsepower ratio , and daytime running lights.

In the past VIN interpretation software has been used mostly by law enforcement, government, divisions of motor vehicles, and other special agencies with authority to use the software. Today, VIN interpretation software is becoming more available for widespread use including availability on the Internet. Consequently, the opportunity to develop innovative uses of vin structures in recording, retrieving, and analysis of vehicle safety data is available.

Privacy issues may if recorded data on VIN includes the full 17 characters of VIN. The this need to protect the privacy of vehicle owners can be accomplished by innovative logic in event recording devices which can record the VIN structure for positions 1-11 along with other information in positions 12-17 which will replace the information normally found in positions 12-17. By maintaining the full 17 character format, a new check digit can be calculated and placed in position 9 of the vin to once again set a control on the validity of the complete vin although the original vin digits no longer appear in positions 12-17. Also, since VIN interpretation software generally requires 17 characters of input, this approach provides opportunities for innovative uses of VIN software to expand the information available in an analysis of a simple data element composed of 17 characters with the VIN structure in positions 1-10 and a numeric code in positions 11-17 which documents the important aspects of an event which occurred.

**Note:** This approach would require special handling for the cases of VINs from small manufacturers.

For example, 6 characters of event information could be placed in positions 11-17 as follows:

Position 11 – Use a single letter to represent the calendar year in which the event occurred -(use the vin - letter scheme for model year with choices of upper case for years 1980-2000, standard font for years 2001-2009, lower case for years 2010-2030, special font for years 2031-2039

**Note:** This scheme for calendar year of event sets up a pattern that is similar to the standardized coding for vehicle model year in position 10 with features to make use of today's wide range of fonts to keep track of model years and event years within the VIN-like data to be recorded and stored for future analysis.

Position 12-14 - Use the "Julian Day -three digit number for the day on which the event occurred.

**Note:** An innovative alternative here is to use the numbers 1-365 for events which occur between the hours from 12:00 AM to Noon and the numbers 366-730 for events between 12:00 noon and 12PM. This innovation is linked to the scheme below for getting a time reference stored in columns 15-16.

Position 15 - The hour can be placed in position 15 as a single character with either a special font or special characters for the 10<sup>th</sup> and 11<sup>th</sup> hour.

Position 16-17 The minutes can be placed here as two digits.

**Note:** The basic idea here is to accurately capture the time elements of the event within the VIN-like data structure. This approach could provide more accuracy in the storage and use of time than current time

recording practices.

To summarize this VIN privatization-event data loading scheme, consider the following example vehicle which has the following in positions 1-10 of the VIN: 1G6KY549YU

Vin interpretation software will reveal that this vehicle is a model year 2000 Cadillac STS 4 Door Sedan equipped with Manual belts, Driver and passenger air bags (front and side) and the engine is a 4.6Litre V8 with multi-port fuel injection.

Now let us assume that this vehicle has been equipped with an on-board event recording device that has been engineered to provide the outputs which includes the VIN-like output event data scheme described above. Consider the fact that this vehicle was involved in a serious crash with another vehicle on November 7,1999 at 4:57 PM. and that the other vehicle was an older vehicle position 1-10 vin characters 9BWBA0307J. This other vehicle was a Volkswagen Fox 2door with active belts only. In this case there were 4 occupants in the Cadillac and only the driver in the Volkswagen Fox. None of the occupants of the Cadillac received injuries but the driver of the Volkswagen was fatally injured. The Fatality Analysis Reporting System( FARS) record shows that both vehicles were moving at 55 mph and the Fox was struck at an angle by the front of the Cadillac at the passenger door of the Fox. The front air bags of the Cadillac both deployed and all occupants of the Cadillac were using shoulder and lap belts. The driver of the Fox was not using belts and was partially ejected.

Using the scheme described above we now have the following two VIN-like data items which contain imbedded information on the time elements of the crash event :

1G6KY5459YX676456 9BWBA0317JX676456

These VIN-like data streams can be subjected to Vin interpretation software to obtain information about the general characteristics of these vehicles. The string X676456 can serve as a time reference that these two vehicle were involved in a crash. This innovative way to display the time elements of a crashed vehicle could be incorporated into the crash event recording equipment as a standard way to keep track of this information and the vehicle which experienced the event. The advantage of such an innovation is that Vin interpretation software can be readily used to discover what vehicle is represented by this record of an event.

Similar information packing schemes can be used in conjunction with the first 8 characters of the vin

structure to facilitate efficient computer handling of masses of vehicles crash information. For example, in this case FARS variables for travel speed, initial point of impact, principal point of impact, and vehicle role are stored as numeric codes which are under format control. Packing this information into a 5 digit code can result in a simplified form of representing these characteristics of the crash as follows:

#### FARS Information (From Format Controls )

Innovative Coding	Travel Speed	Impact1	Impact2	Impacts
55332	55mph	Clock 3	Clock 3	Struck
55^1	55 mph	Clock12	Clock 12	Striking

In this case the special symbol ^ is adapted to represent the frontal contact point for the Cadillac. Thus the combination of characters ^^1 can be used to identify all vehicles which had the role of a striking vehicle. As before these innovative codes can be placed in the VIN-like format to create a way to store data either after a crash occurs or at the time the crash occurs. It is possible that future in-vehicle recording devices will have the capability to generate this kind of information automatically at the time of the crash.

Information about occupants of vehicles can also be condensed into character strings. Information on the 5 occupants in 1999 FARS case #55055 is

Innovative Code	Per. No	Per. Type	Injury Severity	Seating Position	Restrain Use	Air Bag
1410*	1	1	4	1	0	*
10131	1	1	0	1	3	1
20331	2	2	0	3	3	1
2043*	3	2	0	4	3	*
2063*	4	2	0	6	3	*

Note: The FARS two digit code for seating position has been transformed to a one digit code for efficiency setting up the innovative code.

In this example, the innovative code of 1 is created to give an indication that an air bag deployed in the crash. The symbol \* was created when the air bag did not deploy.

Although these innovative codes for the occupant injury levels of the crash relative to the use and deployment of occupant protection devices comes from investigative reports, some of this type of information is expected to be available from on-vehicle crash recorders in the future. Information on person type, seating position, restraint use and air-bag deployment is likely to be available on future vehicles and ultimately analysis of data from these

types of units coupled with actual crash outcomes can lead to reliable ways to project the injury severity levels that are likely in a crash. Capturing this information into a VIN-like data element could facilitate the handling and use of this type of information both at the time of the crash for emergency medical service as well as later in post crash analysis of information coming from crash event recorders as compared to investigative records of the crash outcomes. By keeping the Vin information as a **common link** in all recorded and stored data with other important data elements such as time, crash conditions, etc. in standardized numeric layouts that can be imbedded in a Vin-like structure, the potential for better quality data is enhanced and access to important vehicle characteristics is provided by vin interpretation software and other important linking references.

## STATISTICAL CONSIDERATIONS

A general approach to evaluation of crash data sources is to focus on the severity levels for occupants of vehicles. This approach often leads to count data for ordered categories. In this situation, a statistical method of analysis called RIDIT analysis gives a result which is the probability that a vehicle having the “worst” characteristics will experience higher levels of severity than a vehicle which follows the “best” conditions. For example, using the severity levels in FARS 1999 data we can focus on counts by severity levels relative to air bag information as follows:

Severity Levels	Air Bag Information		
	N/A	Not Deployed	Deployed
A-None	14875	1000	3842
B-Possible Injury	5782	1036	1051
C-Non-Incapacitating	10125	2240	1526
D-Incapacitating	11558	2384	1467
E-Fatal Injury	26106	5255	4348
U-Unknown	218	42	27

With frequency distributions of ordered categories such as this, the statistical method called RIDIT (Relative to an Identified Distribution) Analysis can be applied to compare the ordered severity level counts for one group of vehicles to the ordered severity levels for another group. For example, the severity levels for all vehicles for which restraint system usage and air bag deployment were following the “best” crash worthiness conditions can be compared to a group of vehicles for

which “less than the best” conditions were present.

Innovative techniques can be used to organize vehicles into groups according to conditions that are revealed in crash files. The idea is to create an innovative code and to embed it in the vehicle VIN structure in order to set up opportunities to apply RIDIT analysis concepts for a variety of situations where the first part of the VIN carries information about the vehicle involved and the embedded code carries information about what happened to occupants of the vehicle in a crash situation. For example, seven items of information from the FARS-Person level files can be organized into a 7 digit code representing the following:

Person Type  
Injury Severity Level  
Seat Position  
Restraint Usage  
Air Bag Information  
Ejection  
Ejection Path

When the information for all vehicle occupants is linked to the Vehicle Vin information the resulting VIN-like character strings can be subjected to the RIDIT analysis setups to compare groups of vehicles to one another to reflect on the data that appears in the FARS and other crash files. This type of approach emulates the kind of data that has been projected for future automatic recording of crash events.

## RIDIT ANALYSIS CONCEPTS

RIDIT Analysis is a method for comparing qualitatively ordered categories of count data for one or more sets of observed frequency distributions. One of the sets is chosen as a reference and the distribution for this set is referred to as the “Identified Distribution”. The letters “I” and “D” in RIDIT refer to the identified distribution and the letter “R” stand for “Relative to the Reference set”. The letters “IT” at the end are indicative of the idea of a data transformation analogous to “logit” and “probit” analysis.

The details of the data transformation can be illustrated by an example. Consider a situation were 59 fatal crashes involving 151 vehicle occupants resulted in and observed distribution over the ordered categories of injury severity levels as follow:

**Observed Severity Level Distribution  
(For Reference Situation)**  
**OrderedCategories    ObservedFrequency**

Severity Unknown	2
No Injury	28
Possible Injury	15
Nonincapacitating	21
Incapacitating	26
Fatal	59

#### Ridit Calculations for Reference Set

The reference distribution is transformed by a series of manipulations involving 5 columns of numbers as follows:

Column 1 - The Identified Distribution Reference

Column 2- One-half of entries in column 1

Column 3 - The Cumulative totals of column 1 displaced one category downward (with 0 (zero) on top )

Column 4- Column 2 added to Column 3

Column 5 -Each entry of Column 4 is divided by the Grand total of entries from column 1

Note: Column 5 contains the “ridits” based on the reference counts in column 1

Col 1	Col2	Col3	Col4	Col 5
2	1	0	1.0	0.007
28	14	2	16.0	0.106
15	7.5	30	37.5	0.248
21	10.5	45	55.5	0.368
26	13.0	66	79.0	0.523
59	29.5	92	121.5	0.805

The “ridits” in column 5 can now be used to compare the behavior of a separate set of ordered counts relative to this reference distribution.

Consider the following:

#### Observed Severity Level Distribution (For Comparison Situation)

#### OrderedCategories ObservedFrequency

The test for significance of the difference above can be performed as a standard normal two-tailed test. The standard error of the mean ridit is approximated by the following:

$$\text{Standard Error of Mean Ridit} = \frac{1}{2\sqrt{3 \times 1194}} = .0084$$

Severity Unknown	4
No Injury	100
Possible Injury	103
Nonincapacitating	224
Incapacitating	238
Fatal	525

The procedure for making the comparison is to calculate a single value called the “mean ridit” for the comparison situation. First the “ridits” from the reference group are multiplied by the frequencies of the comparison group. Then these products are added up over all of the ordered categories. Finally the sum of the products is divided by the grand total of the comparison group frequencies. This ratio is the “mean ridit” for the comparison. Thus we have:

Ridits	X	Comparison	=	Products
.007	X	4	=	.028
.106	X	100	=	10.600
.248	X	103	=	25.544
.368	X	224	=	84.432
.523	X	238	=	124.474
.805	X	525	=	422.625

**Total 1194**

**Total of Products= 667. 703**

**Mean Ridit=667.703/1194= 0.56**

The interpretation of this mean ridit is that it is the probability that a person from the comparison group will sustain a more serious injury than a person from the reference group when involved in a serious crash. It is important to note that the mean ridit for the reference group is always 0.50 thus the basis of comparison is the difference between the mean ridit of the comparison group and the mean ridit of the reference group. In this case the difference is **.56 - .50 = .06** and we must now consider the question of whether or not this difference is significant in the statistical sense.

The calculated value **.06/.0084 =7.14** is then compared to the standard normal table with the conclusion that the difference of .06 in this case is significant .

Additional frequency distributions for comparison groups can be examined in this manner to see how they compare to the reference group.

Additional analysis can also be performed to compare the relative behavior of comparison groups to one another. First consider a second comparison group mean ridit calculation as follows:

Ridits	X	Comparison	=	Products
.007	X	2	=	<b>.014</b>
.106	X	384	=	<b>40.704</b>
.248	X	105	=	<b>26.040</b>
.368	X	152	=	<b>55.936</b>
.523	X	147	=	<b>76.881</b>
.805	X	435	=	<b>350.175</b>
<b>Total</b>				<b>1225</b>
<b>Total of Products</b>				<b>=549.75</b>

$$\text{Mean Ridit} = 549.75 / 1225 = 0.45$$

Thus the mean ridit for this comparison group exhibits a lower probability for a member of the comparison group to sustaining a more serious injury in a serious crash than a similar member of the reference group. The difference in mean ridits for this comparison group as compared to the other comparison group is  $0.56 - 0.45 = 0.11$

The standard error for this difference is

$$\begin{aligned} \text{Standard Error of Difference in Mean Ridits} &= \frac{\sqrt{1194} + 1225}{2.0 \times \sqrt{3.0 \times (1984) \times (1225)}} \\ &= .0091 \end{aligned}$$

The calculated value  $.11/.0091 = 12.09$  is compared to the standard normal table to decide the significance of this difference in mean ridits.

In general, RIDIT analysis concepts can be useful for comparing severity levels for injuries to occupants for a variety of circumstances.

## FUTURE DATA SOURCES

As advanced vehicle technologies are applied to better protect occupants in severe crashes, the structural properties of incoming ordered severity information from serious crashes will accumulate in crash files. In many future crashes it is expected that the effect of advanced in-vehicle technologies will be reflected in shifts toward lower severity levels. In many instances the new data will not become a part of the Fatality Reporting System files because of the lack of fatal injuries. Other crash files will contain new information that could be linked to FARS past data to perform the types of Ridit analyses outlined above to estimate and compare the consequences of uses of advanced in-

vehicle technologies in terms of injury severity levels.

In addition, to changes in count data for numbers of events and severity levels, other data quality changes are expected with respect to details such as the exact time of the crash and the exact elapsed time between the time of crash and the time of emergency medical responses to the crash, the exact conditions of restraint usage and air bag deployments.

With these new future data structures, innovative linking schemes, creative uses of data analysis software and information retrieval will be found and applied to better understand the important safety benefits that can result from advanced vehicle technologies [8].

## CONCLUSIONS

Approaches to organizing, linking, and summarizing Safety Assessment data were discussed in terms of moving toward standard ways to collect and store information elements to improve accessibility and quality of data sources.

A statistical method for comparing ordered lists of severity counts was outlined and illustrated. In general a statistical result for an engineering investigation provides valuable insight and information for making decisions leading to improved safety. Only through the collective accumulation of research findings can objective assessments of phenomena under investigation be made.

Innovation and careful consideration of statistical design and analysis of experiments are necessary to bring new technologies and observational data sources into focus so that sound principles of analysis can be easily applied [7,10,11,12]. Standardization of data source elements and design of experiments are necessary to provide the best analysis results.

In this paper the innovation of using one part of a Vehicle Identification Number (VIN-positions 1-10) to identify vehicles and their characteristics and the other part (positions 11-17) to capture crash event conditions such as time, or air bag sensing conditions at the time of crash has been suggested as an advanced way to accurately record and store important information from a crash. This approach is oriented to reliance on the capability of advanced electronic in-vehicle equipment to supply accurate facts at the time of crash as well as on standardized ways to process and assemble vehicle information thru the use of VIN interpretation software. The expectation from applying this type of innovation is better quality data for safety assessments and more flexibility in setting up and applying statistical procedures such as the Ridit Analysis methods.

## REFERENCES

1. Kullgren, Anders, "Validity and Reliability of Vehicle Collision Data", Department of Clinical Neuroscience Section for Personal Injury Prevention, Stockholm 1998
2. Garthe, Elizabeth A. , Mango, Nicholas K. and Prenny, Brad, " A Regional Review of Air Medical Transports for Fatal Motor Vehicle Crashes", Air Medical Journal, Volume 19, Number 3 July-September 2000
3. Donnelly, Bruce R., Schabel, David, Blatt, A.J. Carter, Arthur, "The Automated Collision Notification System", Proceedings- International Symposium on Transportation Recorders, Pages 125-138 , May 3- 5, 1999
- 4 . Scaman, Jeffrey , "The Benefits of Vehicle-Mounted Video Recording Systems", Proceedings- International Symposium on Transportation Recorders Pages 295-300, May 3-5, 1999
5. Mackey, John, J.,Brogan, C.J., Bates, E, Ingalls, S. Howlett, J. , "Digital Eye-Witness Systems", Proceedings-International Symposium on Transportation Recorders, Pages 271-283
6. Kowalick, Thomas Michael, "Proactive Use of Highway Recorded Data via an Event Data Recorder (EDR) to Achieve Nationwide Seat Belt Usage in the 90 th Percentile by 2002", Proceedings-International Symposium on Transportation Recorders, Pages 173-198
7. Washington, Simon, "Conducting Statistical Tests of Hypotheses Five Common Misconceptions Found in Transportation Research", Transportation Research Record , No. 1665
8. U.S. Congress, Office of Technology Assesment, "Advanced Automotive Technology: Visions of a Super-Efficient Family Car", OTA-ETI-638 (Washington, DC: Government Printing Office, September 1995)
9. U.S. Congress, Office of Technology Assesment, "Electronic Surveillance in a Digital Age, OTA-BP-ITC-149 (Washington, DC: U.S. Government Printing Office, July 1995)
10. Glenberg, A.M., "Learning from Data, Introduction To Statistical Reasoning", Lawrence Erlbaum Associates, Mahwah, N.J., 1996
11. Abelson, R.P., "Statistics as Principled Argument", Lawrence Erlbaum Associates, Mahwah, N.J, 1995
12. Harlow, L.L., Malaik, S., and Steiger, "What if There Were No Significance Tests ?", Lawrence Erlbaum Associates, Mahwah, N.J. 1997
13. Bross, Irwin D., "How To Use Redit Analysis" Accident Analysis, Vol. No. ,
14. Tukey, John W., Exploratory Data Analysis", Princeton University, 1977